Robinson LJ, Gray JM, Burt M, Ferrier IN, Gallagher P. Processing of Facial Emotion in Bipolar Depression and Euthymia. *Journal of the International Neuropsychological Society* 2015, 21(9), 709-721.

DOI link to article:

http://dx.doi.org/10.1017/S1355617715000909

Date deposited:

23/10/2015

Embargo release date:

19 April 2016

Processing of Facial Emotion in Bipolar Depression and Euthymia

Lucy J. Robinson [1]

John M. Gray [2]

Mike Burt [3]

I. Nicol Ferrier [2]

Peter Gallagher [2,4]

Published as: **Robinson LJ, Gray JM, Burt M, Ferrier IN, Gallagher P** (2015). Processing of Facial Emotion in Bipolar Depression and Euthymia. *Journal of the International Neuropsychological Society* **21**, 709-721.

[1] School of Psychology, Newcastle University, UK.

[2] Institute of Neuroscience (Academic Psychiatry), Newcastle University, UK.

[3] Department of Psychology, University of Durham, UK.

[4] Newcastle University Institute for Ageing, Newcastle University, UK.

Corresponding Author:

Dr Peter Gallagher, Institute of Neuroscience & Newcastle University Institute for Ageing,

Henry Wellcome Building , Framlington Place, Newcastle upon Tyne, NE2 4HH, UK

Tel.: +44 (0)191 208 7166 Fax.: +44 (0)191 208 5227

e-mail: peter.gallagher@ncl.ac.uk

Word count – Abstract = 249
Word count – Manuscript = 4,996 (excluding title page, abstract, references, tables, and figures)

ABSTRACT

Objective: Previous studies of facial emotion processing in bipolar disorder (BD) have reported conflicting findings. In independently-conducted studies, we investigate facial emotion labelling in euthymic and depressed BD patients using tasks with static and dynamically-morphed images of different emotions displayed at different intensities.

Method: Study 1 included 38 euthymic BD patients and 28 controls. Participants completed two tasks: labelling of static images of basic facial emotions (anger, disgust, fear, happy, sad) shown at different expression intensities; the Eyes Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), which involves recognition of complex emotions using only the eye region of the face. Study 2 included 53 depressed BD patients and 47 controls. Participants completed two tasks: labelling of 'dynamic' facial expressions of the same five basic emotions; the Emotional Hexagon test (A. Young, Perret, Calder, Sprengelmeyer, & Ekman, 2002).

Results: There were no significant group differences on any measures of emotion perception/labelling, compared to controls. A significant group by intensity interaction was observed in both emotion labelling tasks (euthymia and depression), although this effect did not survive the addition of measures of executive function/psychomotor speed as covariates. Only 2.6-15.8% of euthymic patients and 7.8-13.7% of depressed patients scored below the 10^{th} percentile of the controls for total emotion recognition accuracy.

Conclusions: There was no evidence of specific deficits in facial emotion labelling in euthymic or depressed BD patients. Methodological variations – including mood state, sample size, and the cognitive demands of the tasks – may contribute significantly to the variability in findings between studies.

Keywords: Affective disorder, facial emotion labelling, Eyes test, facial expression recognition, emotional hexagon, mood

INTRODUCTION

Bipolar disorder (BD) is a chronic cyclical mood disorder involving periods of elevated (mania/hypomania) and periods of depressed mood. Prospective longitudinal studies have indicated that patients experience mood symptoms around half of the time they have the disorder, but while the characteristic feature of the disorder is (hypo)mania, it is depressive symptoms that are far more prevalent (Judd et al., 2002; Judd et al, 2003). Its aetiology is unknown and a large amount of work in recent years has been undertaken to characterise the functional, cognitive and social deficits associated with the illness (Bonnín et al.; Fagiolini et al., 2005; Goetz, Tohen, Reed, Lorenzo, & Vieta, 2007; Green, Cahill, & Malhi, 2007; MacQueen, Young, & Joffe, 2001; Van Rheenen & Rossell, 2014b). Emotion processing in BD has received increasing attention in an attempt to understand whether some element of dysfunction in the processing of emotional stimuli plays a part in clinical mood symptoms (Van Rheenen & Rossell, 2013). Part of that endeavour has involved exploring facial expression recognition to capture emotion-decoding and labelling processes. Given the central importance of emotional expressions in day-to-day communication, deficits (reduced accuracy) or biases (greater sensitivity to specific emotions or a tendency to consistently interpret emotional stimuli in a particular way) in emotion processing could be of relevance in the experience of mood episodes or in the impaired social functioning seen in BD (Miklowitz, 2011; Sanchez-Moreno et al., 2009).

The findings of studies exploring facial emotion processing in BD are characterised by variability[1] rather than supporting a single deficit or bias in emotion processing (Kohler, Hoffman, Eastman, Healey, & Moberg, 2011; Van Rheenen & Rossell, 2013). In part this may be due to the differences in methods used (e.g. facial image sets, emotion categories used/contrasted with one another, labelling versus discrimination tasks, stimulus display time,

[1] Here we are focussing specifically and exclusively on those studies which involve the requirement to accurately label/name the emotion displayed by the face.

response format), the population studied (or pooled BD subtypes/samples) and sample size. Even in samples of euthymic BD patients there is considerable variability in the findings and conclusions of extant studies, with some reporting specific differences in the recognition of particular emotions, for example, enhanced recognition of disgust (Harmer, Grayson, & Goodwin, 2002), poorer recognition of fear (Martino, Strejilevich, Fassi, Marengo, & Igoa, 2011; Vederman et al., 2012; Venn et al., 2004), poorer recognition of sadness (Vederman, et al., 2012), or poorer recognition of happiness and disgust(Yalcin-Siedentopf et al., 2014); others reporting difficulties with emotion discrimination generally (Addington & Addington, 1998; Bozikas, Tonia, Fokas, Karavatos, & Kosmidis, 2006); and others reporting no significant differences in facial expression recognition (Addington & Addington, 1998; Lee et al., 2013; Lembke & Ketter, 2002; Rowland et al., 2012); or none specifically in patients without a history of psychotic illness features (Thaler et al., 2013). In 'symptomatic' patients the picture is no clearer with some studies reporting no differences on one or other of: recognition, discrimination or sensitivity (Bellack, Blanchard, & Mueser, 1996; Edwards, Pattison, Jackson, & Wales, 2001; Summers, Papadopoulou, Bruno, Cipolotti, & Ron, 2006; Vaskinn et al., 2007). Here clinical heterogeneity is also an issue, with these three studies, respectively, including patients defined as being: generally symptomatic (without specific depression or mania ratings), having 'affective psychosis' (including some patients in mixed and manic states), and a sub-group with varying degrees of residual depressive symptoms. Others have reported differences in recognition in manic patients (generally without exploring specific emotions (Getz, Shear, & Strakowski, 2003), or worse recognition of surprise, but better recognition of disgust in patients compared to controls (Summers, et al., 2006)). In bipolar depression, in two relatively small samples (n=14 and n=21 respectively), differences specifically in sensitivity (i.e. the 'amount' of any particular emotion that needs to be present for the emotion to be correctly recognised) have been reported (Gray et al., 2006; Schaefer, Baumann, Rich, Luckenbaugh, & Zarate, 2010).

To make sense of the disparate and contradictory findings, further studies are needed to develop our understanding of the extent to which emotional processing (specifically the processing and accurate labelling of different facial emotions) may be affected in BD. Studies in relatively large samples of well-characterised patients in clearly-defined mood states and assessing alternative emotion processing/labelling paradigms would go some way to address this gap.

In a recent article in this journal, Van Rheenen and Rossell (2014a) used a series of face-processing paradigms in a pooled sample of patients with BD in different mood states. In the study, three tasks were administered that each employed four basic emotions (happy, sad, anger and fear): emotion labelling of full-intensity dynamically-morphed images (i.e. where static faces are presented rapidly through successive frames from a neutral to the final emotional expression, thereby being perceived as a moving image); emotion labelling of static images of different emotion intensities (high[100%], medium[75%], and low[50%]); and emotion discrimination of static images using the same three intensity levels. When assessing all 3 tasks simultaneously, patients with BD were significantly less accurate than controls generally, although the effect was not seen for all of the tasks when analysed individually. However significant differences between groups on individual emotions were not evident. This led the authors to conclude that there was evidence of a broad deficit in aspects of emotion processing in BD, with effect sizes in the small to medium range. The comprehensive set of tasks used is undoubtedly a strength of the study and serves to highlight the extent to which methodological variations in task demands may contribute to the varied findings in this field. The patient cohort included a mix of depressed, hypomanic, mixed and euthymic states, which were pooled for the primary analyses. While follow-up analyses indicated no statistical differences were reported between these different mood states, the size of the subgroups and complexity of the analyses in a repeated measures design may have impacted on the statistical power of *post hoc* contrasts to detect differences, which the

authors identify as relatively subtle in the group as a whole and which were not detected in all tasks (Van Rheenen & Rossell, 2014a).

In order to further explore the impact of current mood episode and task variations on emotion processing deficits in individuals with BD compared to healthy unaffected controls, the present investigation reports data from two independent studies designed to explore the labelling of facial emotion between bipolar patients and healthy controls, utilising a series of tasks all designed to assess the perception/labelling of emotion from the human face. The first study was conducted in a well-characterised sample of prospectively-verified euthymic BD patients and involved emotion labelling of static images of five basic emotions (angry, happy, fearful, sad, disgusted) at different intensity levels and static facial expression recognition of complex emotions. The second study was conducted in a well-characterised sample of depressed bipolar patients, where it was anticipated that any group differences that resulted from emotion processing deficits would be larger as patients were symptomatic (effectively 'adding' state-related effects to the purported trait-related deficit). To maximise ecological validity of the second study, the tasks involved emotion-labelling of dynamic facial expressions (of the same five basic emotions used in the first study) displayed up to 4 different intensity levels, in addition to a standardised task of processing more ambiguous expression - labelling static images of 'blends' emotions(A. Young, et al., 2002). It was anticipated that emotion labelling deficits would be observed in euthymic patients compared to controls and that between group differences would be significantly greater in symptomatic patients.

STUDY 1: EUTHYMIA

In order to assess the mood-state independence of basic emotion recognition ability in bipolar disorder, study one focussed on testing patients when euthymic.

METHODS

Participants

Sixty-four participants were recruited (n=38 bipolar patients and n=28 controls). Patients were recruited from secondary and tertiary psychiatric services throughout the North East of England. Inclusion criteria comprised: aged 18-65yrs, a DSM-IV SCID diagnosis of bipolar disorder (confirmed using the Structured Clinical Interview (First, Spitzer, Williams, & Gibbon, 1995)) and currently euthymic (≤7 on the 17-item Hamilton Depression Rating Scale (Hamilton, 1960) and the Young Mania Rating Scale (R. Young, Biggs, Ziegler, & Meyer, 1978) which was prospectively verified for 4wks before testing (for details see (Thompson et al., 2005)). Exclusion criteria comprised, current alcohol misuse/dependence, history of head injury with loss of consciousness, neurological illness/major medical illness, ECT within the last 6 months, learning disability or difficulty with fluent use of English language. Patients were not excluded for use of psychotropic medication or for comorbid anxiety disorders (comorbidities were assessed using the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998)).

Control participants were recruited via local advertisements. They were subject to the same exclusion criteria as the patient sample with the addition of no personal history of psychiatric illness and no family history of BD in a first-degree relative. The study was approved by the Newcastle Research Ethics Committee. All participants gave written informed consent.

{TABLE 1 about here}

Measures

Facial Expression Recognition Task – Static Images (FERT-static)

The task used was based on versions used in earlier studies (Harmer et al., 2002, Montagne et al., 2007). Participants were presented with a black and white still facial photograph of a person showing one-of-five facial expressions (angry, disgusted, fearful,

happy, or sad) or neutral. The images used were drawn from the Ekman series (Ekman and Friesen, 1976) and were morphed with neutral (Tiddeman, Burt, & Perrett, 2001) to produce expressions which varied in intensity before being masked from the bottom of the chin to the top of the forehead (thereby covering the hair and ears). Four different individuals were used from the Ekman series (2 male, 2 female) each posing the five expressions plus neutral. This meant each of the expressions was shown sixteen times: four times at each of four intensity levels (20%, 40%, 60%, 80%) (5 emotions x16 presentations=80 stimuli). The neutral expression was shown four times (once per individual,84 stimuli in total).

The picture of the face was presented on a black background (333x482 pixels) on the left hand side of the screen for 1000ms (see Figure 1a). After it had displayed, a solid black mask covered the image and the participant was instructed to indicate the expression (see Figure 1b). The words 'Angry', 'Disgusted', 'Fearful', 'Happy', 'Sad' and 'Neutral' were presented on the side of the screen, listed in alphabetical order. It was not possible for a response to be given when the face was still being displayed.

In order for participants to familiarise themselves with the task, it began with practice trials. This involved six presentations of 100% intensity of each of the 5 emotions and one neutral face. The pictures were of the same individual, who was not used again in the task. The practice trials were presented in the same fixed order to all participants. The 84 experimental trials were presented in a random order to each participant.

Stimuli were presented using Superlab 4.0 (Cedrus) and responses were recorded using 15" CTX resistive-touchscreen monitor. Responses were self-paced with the next stimulus appearing only after the participant had responded to the previous stimulus. The outcome measure of interest was the number of correct responses at each intensity level for each emotion. Reaction time was not analysed as participants were not instructed to respond as quickly as possible.

{FIGURE 1a/b about here}

'Reading-the-Mind-in-the-Eyes' Test

This task is described in detail by Baron-Cohen, et al. (2001). It was used as a measure to assess identification of complex emotions. Although the task is described as a measure of 'theory of mind', it shares features in common with facial expression recognition paradigms and is interpreted in this way here. Participants are shown a single picture of the eye region of a face presented on an A4 page. The picture is surrounded by four adjectives describing a mental state (e.g. perplexed, horrified, astonished). The participant is instructed to identify which of the words they think best describes what the person in the picture is thinking or feeling and circle their choice on a separate answer sheet. After a single practice, 36 experimental items are completed one after the other in a self-paced manner. Response time is not recorded. The outcome measure of interest was the number of correct responses.

Procedure

The tests were administered as part of a wider battery of neuropsychological tests (Robinson, 2010). Participants completed the FERT-static test before the Eyes Test, with unrelated tasks in between. The whole assessment took ~2hrs and participants were able to take breaks.

Data Analysis

Data were analysed using SPSS v.17.0. A significance level of $p<0.05$ was adopted. Patients and controls were compared using independent samples *t*-tests, χ^2-tests or, for tests that involved multiple levels or repetitions, repeated measures ANOVA. For *t*-tests, when Levene's *F*-test identified instances of unequal variance, corrected *p*-values were reported. Effect sizes were calculated using Cohen's d or partial eta squared in the case of factors from the ANOVA (η_p^2). For Cohen's d, positive effect sizes indicate higher scores by the control

group. The proportion of the patient sample scoring <10th percentile of the control group was calculated for each emotion and each intensity level.

RESULTS

FERT-static

The results of the facial expression recognition task are shown in Table 2. The results of a five(emotions) x four(intensity) x two(patient, control) repeated measures ANOVA indicated that there was no significant main effect of group ($F_{1,64}$=0.59, p=0.45, η_p^2=0.01). There was a significant main effect of emotion ($F_{4,256}$=66.44, p<0.001, η_p^2=0.51) and intensity ($F_{3,192}$=583.77, p<0.001, η_p^2=0.90). Follow-up t-tests indicated the main effect of emotion reflected that happy expressions were significantly more easily recognised than each of the other emotions (all p<0.05) and anger was significantly more poorly recognised than the other emotions (all p<0.05) except sadness (p=0.097). There was a significant group x intensity interaction ($F_{3,192}$=2.96, p=0.034, η_p^2=0.04) but follow-up independent samples t-tests did not indicate a significant difference between the groups at any intensity level (all p>0.084) and so the effect could not be related to particular comparisons. The group x emotion interaction was not significant ($F_{4,256}$=1.13, p=0.34, η_p^2=0.02). The three-way interaction between group, intensity and emotion was not significant ($F_{12,768}$=0.51, p=0.91, η_p^2=0.01). Using an independent samples t-test, there was no significant difference between the two groups for recognition of neutral faces (t_{64}=0.81,p=0.42).

{TABLE 2 about here}

'Reading-the-Mind-in-the-Eyes' Test

There were no significant differences between patients and controls for this task (patient mean (s.d.)=26.69 (4.03), control mean (s.d.)=26.79 (3.5),t_{62}=0.10,p=0.93).

SUMMARY OF STUDY 1

In this well-characterised, prospectively verified sample of euthymic BD patients there were no significant differences in emotion labelling of static facial expressions of (primary or complex) emotions, compared to controls. Images were presented to low intensities (20% and 40%), making the task more difficult (although not at floor level) and therefore more likely to both expose group differences avoid ceiling effects in the control group. Despite this, no statistically significant differences were observed. Small effects were observed ($0.2<d<0.5$) for recognition of angry, disgusted and fearful expressions at the higher intensity levels indicating poorer recognition by the patient sample. There was a small effect size indicating better recognition of happiness at the lowest intensity for the patient group. Thus there may be subtle differences in processing/labelling emotions that may become clearer when patients are symptomatic or when stimuli are more naturalistic or ambiguous.

STUDY 2: DEPRESSION

In a second study, we aimed to examine emotional expression labelling in bipolar patients in a current depressive episode. We also administered a dynamic version of the facial emotion recognition test, an approach which has been suggested to hold many advantages over typical static displays, including increased ecological validity (for a review see Krumhuber, Kappas, & Manstead, 2013). In addition we administered a standardised, well-validated 'static' facial emotion labelling task from the *FEEST* battery (A. Young, et al., 2002).

METHODS

Participants

100 participants (n=53 bipolar patients and n=47 matched controls) were recruited. Recruitment was part of a larger research programme into the effects of glucocorticoid receptor antagonists in bipolar depression, which involved a comprehensive baseline assessment of neuropsychological processing, including emotional processing (Gallagher, Gray, Watson, Young, & Ferrier, 2014) (table 1).

Patients were aged 18-65yrs with a diagnosis of BD, confirmed using the Structured Clinical Interview (SCID; First, et al., 1995), and were recruited from secondary/tertiary care in North-East England. All were out-patients and currently in a SCID-defined depressive episode. Patients were excluded if they met criteria for any other current Axis-I disorder or substance dependence/abuse. All were receiving medication at the time of testing (stable ≥4wks). Healthy controls were recruited by general advertisement. All controls were screened to exclude personal/family history (first-degree) of psychiatric illness, significant medical/neurological illness or history of drug/alcohol abuse. The study was approved by the Newcastle and North Tyneside Local Research Ethics Committee. Written informed consent was obtained from all participants.

Measures

Facial Expression Recognition Task – Dynamic Images (FERT-dynamic)

Similar to the FERT-static, this version of the task uses faces from Ekman and Friesen (1976), cropped to isolate the face. Two male and two female faces were used (sets: *jj, pe, pf, mo*). The program rapidly displays the images (~50ms per image), which change from neutral (0% intensity) to the full prototypical emotion (100% intensity) in 5% steps, producing a dynamic morphing effect. This 1000ms 'stream' can be terminated at any of these steps allowing emotional morphs of 5% increments to be possible. For this study, after a short practice block, 80 trials were randomly administered, divided into 4 blocks, permitting a rest between each block. In total there were 16 trials for each of 5 emotions (happy, sad, anger,

disgust, fear). For each of these emotions, 4 intensity levels were used (30-50-70-100%). Participants make their response by pressing one of the five emotion labels presented on the right of the screen. These are only active after the 'morph' has completed and the face disappears.

Benton Facial Recognition Test (short-form) (Benton, Sivan, Hamsher, Varney, & Spreen, 1983)

The BFRT was administered as a control task to examine general face recognition ability. The short form contains 13 trials (maximum score=27). On each item, participants are presented with a target black and white photograph and are asked to choose the target individual from six faces, presented simultaneously with the target photograph.

Emotional Hexagon test (FEEST)

The Emotional Hexagon test from the FEEST was administered according to the standardized instructions (A. Young, et al., 2002). The test utilises one actor (*jj*) from (Ekman and Friesen (1976)) displaying 6 emotional expressions (happiness, surprise, fear, sadness, disgust, anger). Each emotion is blended with the two it is most often confused with, resulting in blends over five continua: happiness–surprise, surprise–fear, fear–sadness, sadness–disgust, disgust–anger; the final blend from anger–happiness completes the hexagon. Blends are displayed in five different proportions of the two emotions: 90:10%, 70:30%, 50:50%, 30:70%, 10:90%. This results in 30 unique stimuli which are displayed randomly 5 times each over the course of the task, giving a total of 150 experimental trials. Participants make their response by pressing one of the six emotion labels presented along the bottom of the screen, which most closely represents the face they saw.

RESULTS

Two patients did not complete the emotion recognition tasks so results are presented for the remaining 51 who had full valid data.

FERT-dynamic

The results of the facial expression recognition task using dynamic stimuli in depressed patients are shown in Table 3. The results of a five(emotion)x four(intensity)x two(group) repeated measures ANOVA indicated that there was no significant main effect of group ($F_{1,96}$=2.23, p=0.14, η_p^2=0.02). There was a significant main effect of emotion ($F_{4,384}$=76.77, p<0.001, η_p^2=0.44) indicating differences in the accuracy of overall emotional labelling (ranging from happy being the most easily detected; average collapsed across group and intensity=95.9%, and disgust being the most difficult; 58.4%) and a main effect of intensity ($F_{3,288}$=104.30, p<0.001,η^2=0.52), with accuracy increasing with increasing intensity. There was no significant group x emotion interaction ($F_{4,384}$=0.71, p=0.59, η_p^2=0.01) and no interaction between group, intensity and emotion ($F_{12,1152}$=1.15, p=0.31, η_p^2=0.01), although the group x intensity interaction was significant ($F_{3,288}$=2.96, p=0.033, η_p^2=0.03), with patients being worse at 30% compared to controls.

The effect sizes showed a small effect size difference for the recognition of disgust and happiness at the lowest intensity level, indicating poorer recognition by the patients. Small effects were also noted for poorer recognition of fear by the patients at the 30%, 50% and 100% intensity levels. There was a medium effect size (0.5≤d<0.8) again showing poorer performance by the patients for the recognition of anger at the lowest intensity level. These are commensurate with the magnitude of effect sizes noted in euthymic patients, not larger as anticipated. As for the euthymic sample, the majority of the calculated effect sizes were d<0.2.

{TABLE 3 about here}

BFRT

BD patients were significantly poorer than controls on the BFRT (t_{98}=-2.41,p=0.02), although this corresponded to only a 1-point difference in performance (BD: mean=22.8, s.d.=2.32; Controls: mean=23.8, s.d.=1.72).

FEEST

Data from the Emotional Hexagon paradigm (figure 2) were available in a sub-set of 51 participants (26 bipolar depressed patients and 25 controls). The results of a six(emotion: angry-disgusted-fearful-happy-sad-surprized)x two(group: patient, control) repeated measures ANOVA indicated that there was no significant main effect of group ($F_{1,49}$=1.56, p=0.22, η_p^2=0.30) or group x emotion interaction ($F_{5,245}$=0.31, p=0.85, η_p^2=0.01). A significant main effect of emotion was observed ($F_{5,245}$=13.66, p<0.0001, η_p^2=0.22). Pairwise comparisons revealed that overall, while not differing from each other, accuracy for happy and sad faces was significantly higher than for all other emotions. Conversely, while not differing from each other, accuracy for disgusted, angry and fearful faces was significantly lower than all other emotions (p<0.05).

{FIGURE 2 about here}

Exploratory analyses (Study 1 and 2)

Correlations

The relationship between performance in emotional labelling and age, and in patients, length of illness was examined. From study 1, for the FERT-static, in euthymic patients the only significant correlations between length of illness and accuracy were for 40% anger (r_s=-0.39, p=0.02) and 80% happy (r_s=-0.41, p=0.01). Significant correlations with age were found for anger at 40% and 60% (r_s=-0.42, p<0.01 and r_s=-0.403, p=0.01), fear at 60% and 80% (r_s=-0.42, p=0.01 and r_s=-0.37, p=0.02) and happy at 60% (r_s=-0.35, p=0.03). In controls, age was correlated with anger at 60% (r_s=-0.45, p=0.02), fear at 40% and 60% (r_s=-

0.52, p=0.005 and r_s=-0.44, p=0.02), and sad at 20% (r_s=-0.44, p=0.02). No significant correlations were observed with the 'Eyes test'. From Study 2, for the FERT-dynamic, in depressed patients, the only significant correlation was between length of illness and 50% disgust (r_s=-0.31, p=0.04) . In controls, age was correlated negatively with fear at 50% (r_s=-0.32, p=0.03) , 70% (r_s=-0.46, p=0.001) and 100% (r_s=-0.33, p=0.02) , and happy at 30% (r_s=-0.32, p=0.03) . For the Emotional Hexagon, the only significant relationship was a positive correlation between disgust and age in patients (r_s=0.44, p=0.02) . The overall effect of age on FERT performance was examined by ANCOVA. Age was a significant covariate in both the depressed ($F_{1,95}$=4.04, p<0.05,η^2=0.04) and euthymic ($F_{1,95}$=9.03, p<0.01,η^2=0.13) analyses but did not affect the overall significant findings (i.e. the significant main effects of emotion, intensity, and a group by intensity interaction).

Impact of general neuropsychological performance

Alongside the emotional recognition tasks in both studies 1 and 2, a broader battery of neuropsychological tests was administered (see Robinson, 2010; Gallagher et al., 2014). To explore the effect of more general (non-emotion related) cognitive processes on performance, we repeated the analysis of FERT data from studies 1 and 2 with the addition of a covariate (ANCOVA). Two commonly used measures utilised in both studies were the 'FAS' test verbal fluency (assessing executive function) (Benton, et al., 1983) and the digit symbol substitution test (DSST)(Wechsler, 1981); assessing psychomotor/processing speed); these were examined independently. In study 1 (euthymia), both the DSST ($F_{1,63}$=6.94, p=0.01, η_p^2=0.10) and the 'FAS' ($F_{1,62}$=4.06, p<0.05, η_p^2=0.06) were significant covariates (FAS: euthymic patients mean=43.3, sd=11.97, controls mean=48.0, sd=11.97; DSST: euthymic patients mean=48.3, sd=11.87, controls mean=54.4, sd=11.53). Their inclusion did not affect the significant main effects of 'emotion' or 'intensity', however the addition of the 'FAS' rendered the previously observed groupxintensity interaction non-significant

($F_{3,186}$=2.29,p=0.09, η_p^2=0.04). In study 2, only the inclusion of the DSST was significant ($F_{1,95}$=5.61, p=0.02, η_p^2=0.06) which again did not affect the significant main effects of 'emotion' or 'intensity', but rendered the groupxintensity interaction non-significant ($F_{3,285}$=2.41, p=0.07, η_p^2=0.025). (FAS: depressed patients mean=38.2, sd=8.88, controls mean=44.5, sd=10.33; DSST: depressed patients mean=48.0, sd=11.76, controls mean=56.4, sd=11.35).

GENERAL DISCUSSION

There were no significant differences between patient and control groups on any of the emotional expression measures used in the present study, with the exception of a group by intensity interaction in facial emotion labelling. Contrary to expectations, neither overall group differences or emotion-specific differences were observed in symptomatic patients nor with tasks using stimuli that were either more ecologically-valid (in the case of the dynamic FERT) or more ambiguous (i.e. labelling complex emotions or blends of different emotions). The only significant differences observed were either not associated with emotional processing (i.e. matching facial identity in depressed BD) or did not remain once general neuropsychological functioning was accounted for (in the case of the FERT interactions between group and intensity). This differs from the recent findings of van Rheenen & Rossell (2014), where a general deficit in emotion recognition and discrimination was observed (Van Rheenen & Rossell, 2014a). It is worth noting that unlike their study, the present studies did not include measures of emotion discrimination. Nonetheless, van Rheenen & Rossell (2014) noted differences on the emotion recognition measures that were not evident in the present studies on similar tasks (emotion recognition of static or dynamic images displayed at different intensities). Our sample included patients in *either* the euthymic or depressed phase of illness and explored the two groups separately. Combining groups of patients in different

symptomatic states and including patients in the manic or hypomanic state could be one reason why the results differ. A recent meta-analysis focussing exclusively on euthymia reported a significant effect for the Eyes Test in contrast to the present finding, although the effect size of the deficit in patients was small (Hedges' g=0.27) (Samamé, Martino, & Strejilevich, 2015). Interestingly, the labelling of several individual emotions (i.e. anger, sadness, disgust) from facial emotion perception studies was not significantly different between patients and controls (Hedges' g=0.15-0.25), although recognition of surprise and fear was significantly worse but again this effect was small (Hedges' g=0.22-0.29). After excluding one outlying study, significantly greater impairment was observed for disgust and fear recognition in patients (Hedges' g=0.39-0.43).

The relatively comprehensive set of emotion recognition tests, including paradigms that are generally considered more difficult and therefore more likely to expose a deficit or bias (e.g. static images of low-intensity emotions), combined with large samples of well-characterised patient groups are strengths of the present study. As with many studies in patient samples, low statistical power is a concern in the present study. While the present analyses were adequately powered ($1-\beta \geq 80\%$) to identify large effect size differences for main effects of group, power was lower to detect smaller effect sizes, especially from interactions. Indeed, the observed effect sizes indicated small effects ($d<0.2$) on some measures, although many were below this threshold. This study adds to others (Addington & Addington, 1998; Bellack, et al., 1996; Edwards, et al., 2001; Lembke & Ketter, 2002; Rowland, et al., 2012; Vaskinn, et al., 2007) that have not reported evidence of significant impairment in facial emotion recognition in BD. It is difficult to infer directly from statistical effect size to clinical significance, but it seems this element of emotion processing (specifically the labelling of displayed emotion) may be of limited importance in understanding the presentation of those with this disorder. However, we reiterate the specificity of our findings here as we address only one aspect of emotion processing – the

perception and labelling of emotion transmitted by the face/facial features. Numerous other processes have been examined in mood disorders, such as attentional bias for emotional stimuli, go/no-go biases and memory/recall of emotional information (Jongen, Smulders, Ranson, Arts, & Krabbendam, 2007; Rubinsztein, Michael, Underwood, Tempest, & Sahakian, 2006; Wessa & Linke, 2009). Our studies are concerned only with this labelling process and cannot speak to questions around other processes, although it is critical for future studies to determine that their findings are clearly attributable to the *emotional* process per se and not secondary to a more general neurocognitive deficit.

It is important to note that the patient samples in our study did show significant neuropsychological deficits with large effect sizes in many domains of 'cold' cognition (Gallagher, et al., 2014; Robinson, 2010) and therefore the absence of differences is not a consequence of recruiting high-performing patients with BD. To derive a sense of the relative scale of 'impairment', the proportion of the patient group falling below the 5^{th}/10^{th} percentile of controls can be examined (Gallagher, et al., 2014; Thompson, et al., 2005). In the euthymic sample, the proportion of patients scoring below the 10^{th} percentile on cognitive measures (administered alongside the facial expression battery) ranged from 2.6%-53.8% (Robinson, 2010). These tests included measures of executive function, verbal declarative memory, working memory and psychomotor speed. Those domains showing the largest proportion of low-scoring patients were executive measures (category fluency, 53.8%) and verbal declarative memory (list-learning total recall, 42.1%). In contrast, the proportion of patients scoring below the 10^{th} percentile on the facial expression recognition test, after separating by intensity level, ranged from 2.6%-18.4% (for the total this was 2.6-15.8%) suggesting there is less evidence of potential impairment on these measures. Data for the depressed patients showed a similar pattern. The cognitive performance of the depressed sample is detailed elsewhere (Gallagher, et al., 2014): patients performed significantly worse on 18/26 measures examined, with large effect sizes ($d>0.8$) on tests of speed of processing, verbal learning and

specific executive/working memory processes. Almost all tests produced at least one outcome measure on which ~25–50% of the BD sample performed at more than 1s.d. below the control mean. Patients performing below the controls' 10th percentile for measures of accuracy ranged from 11.3%-47.2%. However, in the present study, for the facial expression recognition task, examining the separate intensities this ranged from 0%-29.4% (for the total this was 7.8-13.7%). Importantly, our exploratory analyses showed that by including measures of executive function or psychomotor/processing speed as covariates could account for the group by intensity interactions seen if both FERT tasks. This is in line with a previous study suggesting that deficits in Theory of Mind and emotion labelling may be in part mediated by attention-executive deficits (Martino, et al., 2011). However, several caveats should be noted. First, directly comparing 'cold' and 'hot' cognitive tasks is problematic if the discriminating power of the tasks differ; indeed with tests of differing reliabilities, the measure with the higher reliability coefficient will record a greater performance decrement for less able participants (Chapman & Chapman, 1973). There are also several possible interpretations for the observed interaction effect between group and intensity – while we suggest that the effect is a consequence of generalised deficits leading to difficulties with the most difficult/ambiguous stimuli (i.e. stimuli with the lowest 'information' content) we cannot rule out the possibility that this is reflecting a specific deficit in low-level emotional perception (for both positive and negative emotions). Therefore it is important for future studies to explore this effect within the task design itself (rather than *post hoc* through statistical methods).

Given the extent of these neuropsychological deficits, it might be that where individuals with BD have shown performance deficits on tasks involving facial expression perception previously, some of these findings may have been secondary to general difficulties in performing (lab-based experimental) tasks, rather than deficits in facial expression perception *per se*. However, the effect of such general deficits might be expected to be fairly

small (since one would hope that the assessments of facial expression perception have a good degree of specificity) and emerge as significant in a fairly random fashion in some experiments but not others and, within these experiments, in some conditions but not others (contingent upon the precise demands of the task/condition); this pattern seems to describe the literature reviewed previously in BD. For example, where facial expression perception experiments and analyses overlap with cognitive domains in which individuals with BD have deficits, they would be more likely to report significant results with a greater effect size. It is of interest that in fMRI studies it has been demonstrated that patterns of activation differ according to the demands of the task. Direct matching of emotional facial expressions has been found to increase amygdala activation while the selection of the label that matches (e.g. 'afraid') results in greater right prefrontal cortex activation (Hariri, Bookheimer, & Mazziotta, 2000). Therefore tasks which examine emotion discrimination compared to labelling may be tapping different aspects of processing.

These methodological differences may partially account for some of the variability in findings to date (we refer specifically to the accuracy decrement here, rather than bias). For example, tasks with a response format that have a high memory load, complex instructions or time-pressured responses may be more likely to show group differences. Indeed, the greatest proportion of the depressed sample scoring below the 10^{th} percentile for the control group occurred at the 30% level of intensity, which was the stimulus that was displayed for the shortest amount of time. Future studies should also consider how the specifics of the response format can potentially affect the outcome of studies of this nature. For example, it is important to be mindful that the majority of studies are fixed-choice paradigms (i.e. there is no "don't know" option, such as in standardized measures like the Ekman-60 (A. Young, et al., 2002)). Therefore if stimuli are presented quickly or are ambiguous, participants still have to select one of the options to move to the next trial. Therefore patients (who may simply be slightly slowed in general processing speed or decision making) are more likely to 'miss'

stimuli and select a random response to move on – this is not an emotional processing bias/deficit, although it may seem so if systematic factors influence the response chosen (e.g. the response option closest to the participant's hand). It should also be noted that in tasks of this nature, the majority of the available responses are 'negative' emotions, with 'happy' typically the only overt positive emotion available. Therefore, with regard to the occurrence of this latter phenomena, any form of systematic response bias will lead to a 'deficit' in the perception of one emotion and an increase in another, which will typically be another 'negative' emotion.

A further point to consider is how findings in this area are interpreted. For example, results that have demonstrated reduced accuracy of labelling specific expressions have been interpreted as supporting the notion that emotion perception decrements are evident in BD (Vederman et al., 2012). Other studies have interpreted increased correct recognition of specific emotions (e.g. disgust) as possibly being linked to low self-esteem and other cognitive biases in BD (Harmer et al., 2002). It is therefore important to consider the precise nature of the task demands and the social processes being assessed to avoid a situation in which both increased and decreased accuracy is considered as reflecting a 'negative outcome'. It is also necessary to consider the potential difference between greater accuracy, which may reflect hypersensitivity to characteristic features of emotional expressions and hence 'more accurate' social perception, versus a 'true' bias where stimuli (especially ambiguous stimuli) are consistently interpreted as showing a particular emotion (Leppanen, 2006), suggesting top down influences are affecting the interpretation of incoming information so the individual 'sees' a particular emotion when it may not be present (Martino, et al., 2011). It is worth noting that the two processes (mood related bias and general deficit in accurate responding) may work counter each other in particular cases. Further work is needed to develop an understanding of the circumstances in which accuracy decrements occur and those where hypersensitivity or bias may occur.

There are a number of limitations of the present study to be considered. Firstly, low statistical power for the interaction analyses has already been mentioned. This difficulty is commonly encountered in this area of investigation and is likely to contribute to the varied findings. More widespread reporting of effect sizes alongside inferential statistics would help clarify whether studies are broadly finding group differences of a similar magnitude or, if not, it may help to identify which methodological variations impact most markedly on group differences. Secondly, we did not administer the same tests to both patient groups, which raises the possibility that some measures may have shown differences had both groups received the same tasks. However, three of the tasks used the same image set and similar intensities of emotions and all involved a range of difficulty in the stimuli presented, thereby offering the opportunity for even a subtle deficit to become evident. Also, using the two different experimental expression recognition tests suggests the lack of difference is not specific to a methodological feature of one particular task. Furthermore, the depressed sample were administered standardised measures (e.g. the Emotional Hexagon(A. Young, et al., 2002)) alongside the other tasks and did show pronounced deficits in other aspects of cognitive function. Thirdly, although we utilised a dynamic emotional expression task to increase the ecological validity of the task, some studies have suggested that dynamic facial movements play only a small role in the ability to identify emotion from facial expressions (Gold et al., 2013). Nonetheless, employing different variants of facial emotion stimuli develops our understanding of the robustness or otherwise of any effect irrespective of ecological validity. Recently it has been demonstrated that impairments can be observed in dynamic (videotaped) displays of emotion and more complex aspects of social communication in BD, in the absence of differences in labelling static images of facial emotion (Rowland, et al., 2012). Therefore the use of methodologies that capture the real-world complexities and subtleties of social interaction may prove important tools for future studies to explore emotional processing deficits in BD.

Based upon our current findings and the mixed findings of the literature we conclude there is little evidence of abnormalities in explicit facial emotion identification in euthymic or depressed patients, within the parameters examined in the present studies. Future studies should address the methodological issues in this area of research- especially using paradigms with limited memory load or time pressure- in order to build a more complete picture of emotion processing in BD and how or whether it is of relevance in our understanding of this illness.

References

Addington, J., & Addington, D. (1998). Facial affect recognition and information processing in schizophrenia and bipolar disorder. *Schizophrenia Research, 32*(3), 171-181.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bellack, A. S., Blanchard, J. J., & Mueser, K. T. (1996). Cue Availability and Affect Perception in Schizophrenia. *Schizophrenia Bulletin, 22*(3), 535-544. doi: 10.1093/schbul/22.3.535

Benton, A., Sivan, A., Hamsher, K., Varney, N., & Spreen, O. (1983). *Contribution to neuropsychological assessment*. NY: Oxford University Press.

Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., et al. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: A long-term, follow-up study. *Journal of Affective Disorders, In Press, Corrected Proof*.

Bozikas, V. P., Tonia, T., Fokas, K., Karavatos, A., & Kosmidis, M. H. (2006). Impaired emotion processing in remitted patients with bipolar disorder. *Journal of Affective Disorders, 91*(1), 53-56.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychol Bull., 79*(6), 380-385.

Edwards, J., Pattison, P. E., Jackson, H. J., & Wales, R. J. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophrenia Research, 48*(2–3), 235-253. doi: http://dx.doi.org/10.1016/S0920-9964(00)00099-2

Ekman, P., & Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Fagiolini, A., Kupfer, D. J., Masalehdan, A., Scott, J. A., Houck, P. R., & Frank, E. (2005). Functional impairment in the remission phase of bipolar disorder. *Bipolar Disorders, 7*(3), 281-285.

First, M. B., Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1995). *Structured Clinical Interview for DSM-IV (SCID-IV)*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Gallagher, P., Gray, J. M., Watson, S., Young, A. H., & Ferrier, I. N. (2014). Neurocognitive functioning in bipolar depression: a component structure analysis. *Psychol Med, 44*(5), 961-974. doi: 10.1017/s0033291713001487

Getz, G. E., Shear, P. K., & Strakowski, S. M. (2003). Facial affect recognition deficits in bipolar disorder. *Journal of the International Neuropsychological Society, 9*(4), 623-632.

Goetz, I., Tohen, M., Reed, C., Lorenzo, M., & Vieta, E. (2007). Functional impairment in patients with mania: baseline results of the EMBLEM study. *Bipolar Disorders, 9*(1-2), 45-52. doi: 10.1111/j.1399-5618.2007.00325.x

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., et al. (2013). The efficiency of dynamic and static facial expression recognition. *J Vis, 13*(5). doi: 10.1167/13.5.23

Gray, J., Venn, H., Montagne, B., Murray, L., Burt, M., Frigerio, E., et al. (2006). Bipolar patients show mood-congruent biases in sensitivity to facial expressions of emotion when exhibiting depressed symptoms, but not when exhibiting manic symptoms. *Cogn Neuropsychiatry, 11*(6), 505-520. doi: 10.1080/13546800544000028

Green, M. J., Cahill, C. M., & Malhi, G. S. (2007). The cognitive and neurophysiological basis of emotion dysregulation in bipolar disorder. *Journal of Affective Disorders, 103*, 29-42. doi: 10.1016/jjad.2007.01.024

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology Neurosurgery and Psychiatry, 23*, 56-62.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport, 11*(1), 43-48.

Harmer, C. J., Grayson, L., & Goodwin, G. M. (2002). Enhanced recognition of disgust in bipolar illness. *Biological Psychiatry, 51*(4), 298-304.

Jongen, E. M. M., Smulders, F. T. Y., Ranson, S. M. G., Arts, B. M. G., & Krabbendam, L. (2007). Attentional bias and general orienting processes in bipolar disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(2), 168-183.

Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K., & Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: a quantitative review. *Psychiatry Res, 188*(3), 303-309. doi: 10.1016/j.psychres.2011.04.019

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5*(1), 41-46.

Lee, J., Altshuler, L., Glahn, D. C., Miklowitz, D. J., Ochsner, K., & Green, M. F. (2013). Social and Nonsocial Cognition in Bipolar Disorder and Schizophrenia: Relative Levels of Impairment. *American Journal of Psychiatry, 170*(3), 334-341. doi: doi:10.1176/appi.ajp.2012.12040490

Lembke, A., & Ketter, T. A. (2002). Impaired recognition of facial emotion in mania. *American Journal of Psychiatry, 159*(2), 302-304.

Leppanen, J. M. (2006). Emotional information processing in mood disorders: a review of behavioral and neuroimaging findings. *Current Opinion in Psychiatry, 19*(1), 34-39.

MacQueen, G. M., Young, L. T., & Joffe, R. T. (2001). A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatrica Scandinavica, 103*(3), 163-170.

Martino, D. J., Strejilevich, S. A., Fassi, G., Marengo, E., & Igoa, A. (2011). Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research, 189*(3), 379-384. doi: http://dx.doi.org/10.1016/j.psychres.2011.04.033

Miklowitz, D. J. (2011). Functional impairment, stress, and psychosocial intervention in bipolar disorder. *Curr Psychiatry Rep, 13*(6), 504-512. doi: 10.1007/s11920-011-0227-x

Robinson, L. J. (2010). *Neuropsychological Performance, Emotion Processing and Psychosocial Function in Bipolar Disorder.* PhD, Newcastle University, Newcastle upon Tyne.

Rowland, J. E., Hamilton, M. K., Vella, N., Lino, B. J., Mitchell, P. B., & Green, M. J. (2012). Adaptive Associations between Social Cognition and Emotion Regulation are Absent in Schizophrenia and Bipolar Disorder. *Frontiers in Psychology, 3*, 607.

Rubinsztein, J. S., Michael, A., Underwood, B. R., Tempest, M., & Sahakian, B. J. (2006). Impaired cognition and decision-making in bipolar depression but no 'affective bias' evident. *Psychological Medicine, 36*(5), 629-639.

Samamé, C., Martino, D. J., & Strejilevich, S. A. (2015). An individual task meta-analysis of social cognition in euthymic bipolar disorders. *Journal of Affective Disorders, 173*, 146-153.

Sanchez-Moreno, J., Martinez-Aran, A., Tabares-Seisdedos, R., Torrent, C., Vieta, E., & Ayuso-Mateos, J. L. (2009). Functioning and disability in bipolar disorder: an extensive review. *Psychotherapy & Psychosomatics, 78*(5), 285-297. doi: 10.1159/000228249 [doi]

Schaefer, K. L., Baumann, J., Rich, B. A., Luckenbaugh, D. A., & Zarate, C. A., Jr. (2010). Perception of facial emotion in adults with bipolar or unipolar depression and controls. *J Psychiatr Res, 44*(16), 1229-1235. doi: 10.1016/j.jpsychires.2010.04.024

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*, 22-33; 34-57.

Summers, M., Papadopoulou, K., Bruno, S., Cipolotti, L., & Ron, M. A. (2006). Bipolar I and bipolar II disorder: cognition and emotion processing. *Psychological Medicine, 36*(12), 1799-1809.

Thaler, N. S., Strauss, G. P., Sutton, G. P., Vertinski, M., Ringdahl, E. N., Snyder, J. S., et al. (2013). Emotion perception abnormalities across sensory modalities in bipolar disorder with psychotic features and schizophrenia. *Schizophrenia Research, 147*(2–3), 287-292. doi: http://dx.doi.org/10.1016/j.schres.2013.04.001

Thompson, J. M., Gallagher, P., Hughes, J. H., Watson, S., Gray, J. M., Ferrier, I. N., et al. (2005). Neurocognitive impairment in euthymic patients with bipolar affective disorder. *British Journal of Psychiatry, 186*, 32-40.

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21*(5), 42-50. doi: 10.1109/38.946630

Van Rheenen, T. E., & Rossell, S. L. (2013). Is the non-verbal behavioural emotion-processing profile of bipolar disorder impaired? A critical review. *Acta Psychiatrica Scandinavica, 128*(3), 163-178. doi: 10.1111/acps.12125

Van Rheenen, T. E., & Rossell, S. L. (2014a). Let's face it: facial emotion processing is impaired in bipolar disorder. *J Int Neuropsychol Soc, 20*(2), 200-208. doi: 10.1017/s1355617713001367

Van Rheenen, T. E., & Rossell, S. L. (2014b). Objective and subjective psychosocial functioning in bipolar disorder: An investigation of the relative importance of neurocognition, social cognition and emotion regulation. *Journal of Affective Disorders, 162*(0), 134-141. doi: http://dx.doi.org/10.1016/j.jad.2014.03.043

Vaskinn, A., Sundet, K., Friis, S., Simonsen, C., Birkenaes, A. B., Engh, J. A., et al. (2007). The effect of gender on emotion perception in schizophrenia and bipolar disorder. *Acta Psychiatrica Scandinavica, 116*(4), 263-270. doi: 10.1111/j.1600-0447.2007.00991.x

Vederman, A. C., Weisenbach, S. L., Rapport, L. J., Leon, H. M., Haase, B. D., Franti, L. M., et al. (2012). Modality-specific alterations in the perception of emotional stimuli in Bipolar Disorder compared to Healthy Controls and Major Depressive Disorder. *Cortex, 48*(8), 1027-1034. doi: http://dx.doi.org/10.1016/j.cortex.2011.03.017

Venn, H. R., Gray, J. M., Montagne, B., Murray, L. K., Michael Burt, D., Frigerio, E., et al. (2004). Perception of facial expressions of emotion in bipolar disorder. *Bipolar Disorders, 6*(4), 286-293.

Wechsler, D. (1981). *WAIS-R manual, Wechsler Adult Intelligence Scale-Revised*. Cleveland, OH: Psychological Corporation.

Wessa, M., & Linke, J. (2009). Emotional processing in bipolar disorder: Behavioural and neuroimaging findings. *International Review of Psychiatry, 21*(4), 357-367. doi: 10.1080/09540260902962156

Yalcin-Siedentopf, N., Hoertnagl, C. M., Biedermann, F., Baumgartner, S., Deisenhammer, E. A., Hausmann, A., et al. (2014). Facial affect recognition in symptomatically remitted patients with schizophrenia and bipolar disorder. *Schizophrenia Research, 152*(2–3), 440-445. doi: http://dx.doi.org/10.1016/j.schres.2013.11.024

Young, A., Perret, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: stimuli and tests (FEEST)*. Bury, St. Edmunds: Thames Valley Test Company.

Young, R., Biggs, J., Ziegler, V., & Meyer, D. (1978). A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry, 133*(5), 429-435.

Acknowledgements

This work was supported by grant funding from the Stanley Medical Research Institute (REF: 03T-429) and the Medical Research Council (GU0401207). LJR, PG and INF received Research Capability Funding support from the Northumberland, Tyne and Wear NHS Foundation Trust. We thank the Mental Health Foundation (north-east) for supporting the purchase of test materials.

We are grateful to the participants who contributed to the research and to those clinicians involved in the wider research programme, including recruitment and screening: Stuart Watson, David Watson, Allan Young, Niraj Ahuja, Sankalpa Basu, Jane Carlile, Louise Golightly, Thiyyancheri Harikumar, Patrick Keown, Samer Makhoul, Anuradha Menon, Gavin Mercer, Rajesh Nair, Bruce Owen, and Nanda Palanichamy. We thank Daniel Hedley for assistance with the analysis of data from the FEEST.

The authors have no conflicts of interest to declare in relation to the manuscript.

Figure legends

Figure 1: Screen shot of the response options for the Facial Expression Recognition Task using static images. Faces were presented in the black rectangle for 1 second.

Figure 2: Results of the facial expression recognition task in depressed patients using static blended stimuli (emotional hexagon).

Table 1: Demographic details of the patient samples

	Control		Patient			
	mean	**s.d.**	**mean**	**s.d.**	**t/χ^2**	**p**
Euthymic Group	n=28		n=38			
Demographics						
Age	46.5	10.8	44.8	12.8	0.54	0.592
Male/female						
Male: n (%)	13	(46.4)	17	(44.7)	0.00	0.982
Female: n (%)	15	(53.6)	21	(55.3)		
NART IQ	114.4	8.9	111.2	9.6	1.62	0.110
Years of education	16.8	2.9	15.5	3.8	1.72	0.090
Mood Symptoms						
HDRS-17	-	-	3.8	2.1	-	-
YMRS	-	-	0.7	1.6	-	-
BDI	1.3	1.8	7.2	6.9	-4.76	<0.001
AMRS	2.1	3.3	1.8	3.3	0.44	0.665
Depressed Group	n=47		n=53			
Demographics						
Age	45.0	13.7	47.3	9.6	0.97	0.343
Male/female						
Male: n (%)	28	(59.6)	33	(62.3)	0.08	0.783
Female: n (%)	19	(40.4)	20	(37.7)		
Nart IQ	112.5	11.2	108.9	10.5	1.63	0.107
Years of education	14.4	4.0	14.4	3.2	0.05	0.961
Mood Symptoms						
HDRS-17	-	-	19.7	4.9	-	-
YMRS	-	-	1.5	1.8		
BDI	1.0	1.5	26.0	11.4	10.46	<0.001
AMRS	-	-	-	-	-	-

NART, National Adult Reading Test; HDRS-17, Hamilton Depression Rating Scale 17-item; YMRS, Young Mania Rating Scale; BDI, Beck Depression Inventory; AMRS, Altman Mania Rating Scale

Table 2: Results of the facial expression recognition task in euthymic patients using static stimuli. Means and standard deviations of % correct at each intensity level for each emotion.

	Control (n=28)		Bipolar (n=38)			%BD below 10th percentile[1]
	mean	s.d.	mean	s.d.	d	
Angry						
Correct 20%	9.82	12.43	10.53	14.97	-0.05	0.0
Correct 40%	31.25	22.18	32.24	20.06	-0.05	0.0
Correct 60%	63.39	31.54	50.66	24.31	0.46	2.6
Correct 80%	78.57	24.26	70.39	23.86	0.34	2.6
Correct Total %	45.78	16.66	40.95	13.39	0.33	5.3
Disgust						
Correct 20%	6.25	14.63	5.26	11.85	0.08	0.0
Correct 40%	45.54	24.58	40.13	26.98	0.21	18.4
Correct 60%	76.79	22.49	65.79	29.88	0.41	18.4
Correct 80%	76.79	25.39	71.71	22.64	0.21	7.9
Correct Total %	51.35	14.96	45.73	15.91	0.36	15.8
Fear						
Correct 20%	10.71	15.85	16.45	20.37	-0.31	0.0
Correct 40%	66.07	26.54	63.82	20.71	0.10	0.0
Correct 60%	83.93	20.65	76.97	27.50	0.28	10.5
Correct 80%	84.82	21.88	80.92	21.30	0.18	5.3
Correct Total %	61.40	15.22	59.54	14.73	0.12	7.9
Happy						
Correct 20%	35.71	26.73	46.71	28.58	-0.40	0.0
Correct 40%	85.71	18.54	84.87	21.39	0.04	18.4
Correct 60%	93.75	12.95	95.39	11.41	-0.14	2.6
Correct 80%	95.54	11.89	97.37	7.78	-0.19	0.0
Correct Total %	77.70	12.45	81.09	13.36	-0.26	13.2
Sad						
Correct 20%	16.96	18.07	18.42	18.09	-0.08	0.0
Correct 40%	43.75	26.02	46.05	32.64	-0.08	0.0
Correct 60%	59.82	26.65	60.53	25.09	-0.03	5.3
Correct 80%	66.07	29.04	65.79	23.55	0.01	0.0
Correct Total %	46.67	18.95	47.71	16.86	-0.06	2.6
Neutral						
Correct Total %	81.25	23.20	76.32	25.30	0.20	7.9
Void	0.21	1.13	0.26	0.76	-0.05	0.0

[1] Of the control sample

Table 3: Results of the facial expression recognition task in depressed patients using dynamic stimuli. Means and standard deviations of number correct at each intensity level for each emotion.

	Control (n=47)		Bipolar (n=51)			% BD below 10th percentile[1]
	mean	s.d.	mean	s.d.	d	
Angry						
Correct 30%	50.00	26.58	34.80	27.87	0.56	29.4
Correct 50%	63.30	28.00	58.82	26.84	0.16	5.9
Correct 75%	77.66	23.45	74.51	26.69	0.13	13.7
Correct 100%	81.91	24.84	79.90	25.50	0.08	5.9
Total%	68.22	17.57	62.01	19.28	0.34	9.8
Disgust						
Correct 30%	59.57	24.20	46.08	31.77	0.48	17.6
Correct 50%	55.85	29.59	58.82	33.85	-0.09	9.8
Correct 75%	62.77	28.96	62.25	28.45	0.02	3.9
Correct 100%	61.70	26.50	60.29	31.30	0.05	9.8
Total%	59.97	20.96	56.86	25.07	0.13	13.7
Fear						
Correct 30%	63.83	22.00	59.80	26.02	0.17	17.6
Correct 50%	82.45	20.13	73.53	24.19	0.40	7.8
Correct 75%	83.51	19.70	78.92	17.59	0.25	3.9
Correct 100%	78.72	22.71	74.51	20.30	0.20	2.0
Total%	77.13	16.60	71.69	15.68	0.34	9.8
Happy						
Correct 30%	90.96	16.00	86.76	18.27	0.24	9.8
Correct 50%	96.81	8.43	96.57	10.02	0.03	2.0
Correct 75%	97.87	7.05	98.53	5.94	-0.10	5.9
Correct 100%	100.00	0.00	99.51	3.50	0.19	2.8
Total%	96.41	5.94	95.34	6.71	0.17	9.8
Sad						
Correct 30%	51.60	30.13	52.94	29.00	-0.05	0.0
Correct 50%	59.57	29.28	62.75	29.31	-0.11	5.9
Correct 75%	62.23	26.52	60.29	31.69	0.07	11.8
Correct 100%	70.21	24.80	70.59	24.34	-0.02	0.0
Total%	60.90	21.86	61.64	22.40	-0.03	7.8

[1] Of the control sample

Figure 1

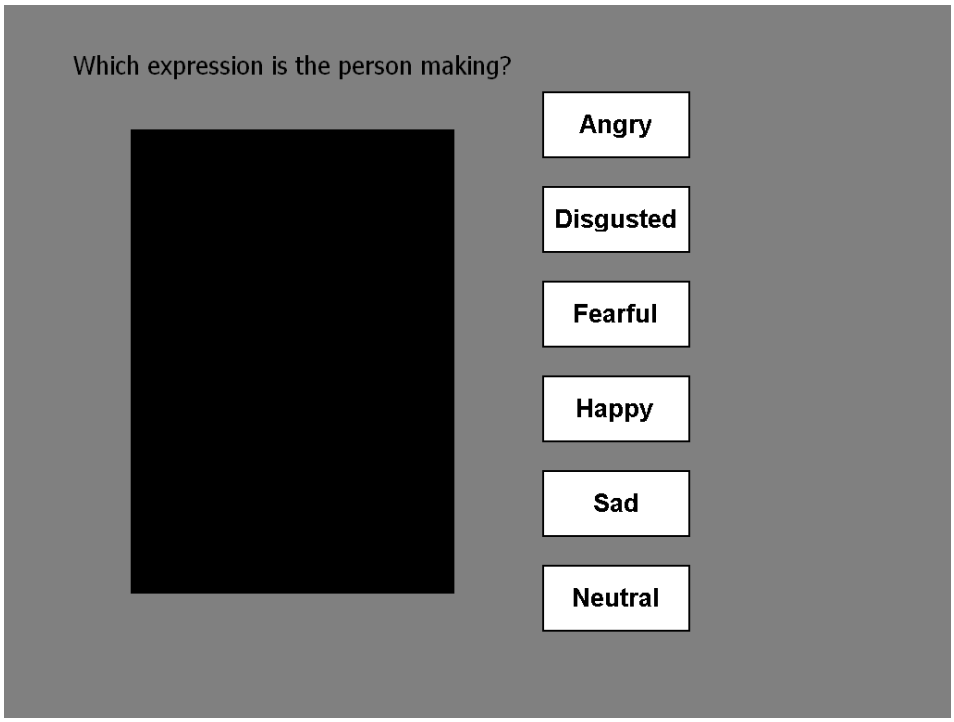
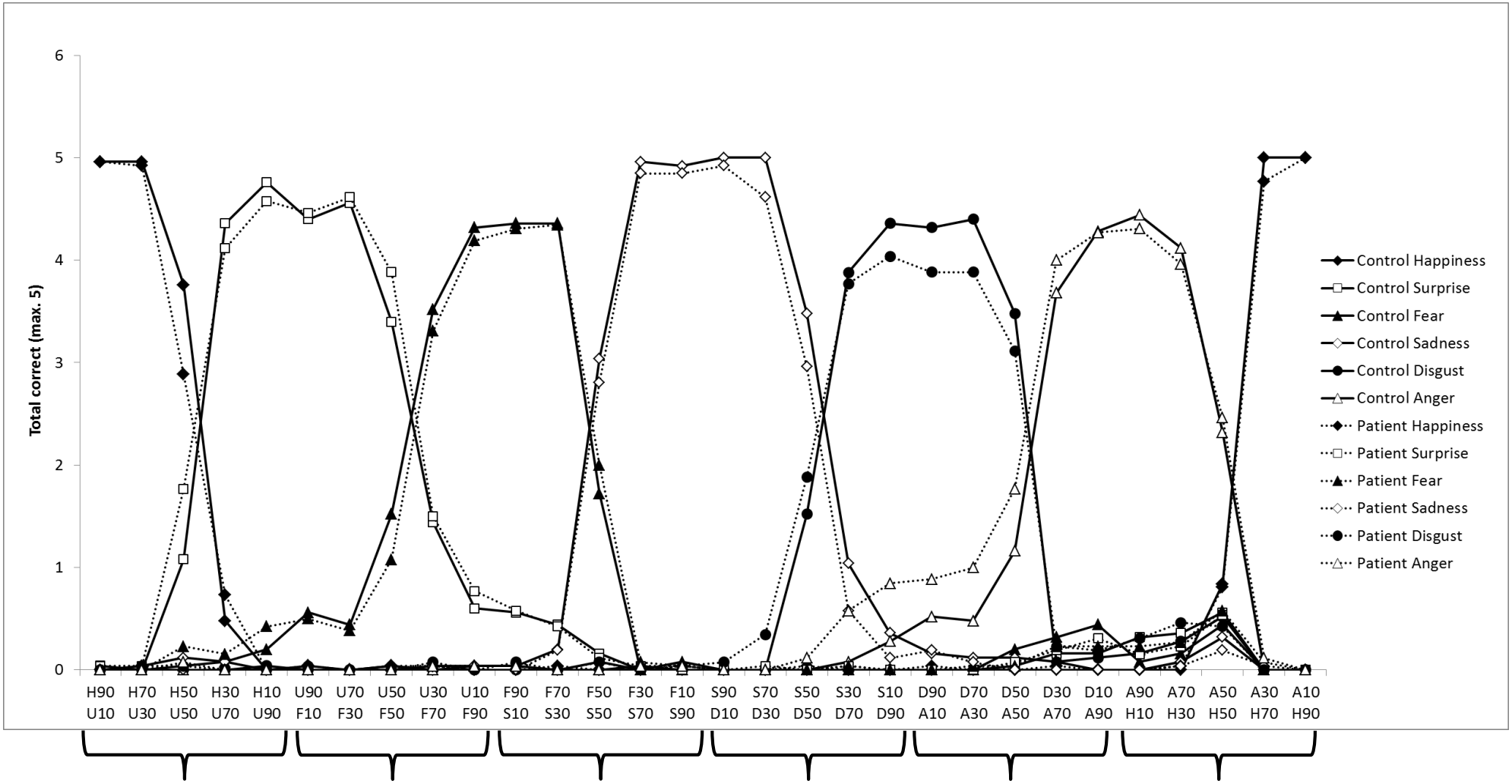

Which expression is the person making?
Angry
Disgusted
Fearful
Happy
Sad
Neutral

Figure 2


Total correct (max. 5)
0
1
2
3
4
5
6
H90 U10
H70 U30
H50 U50
H30 U70
H10 U90
U90 F10
U70 F30
U50 F50
U30 F70
U10 F90
F90 S10
F70 S30
F50 S50
F30 S70
F10 S90
S90 D10
S70 D30
S50 D50
S30 D70
S10 D90
D90 A10
D70 A30
D50 A50
D30 A70
D10 A90
A90 H10
A70 H30
A50 H50
A30 H70
A10 H90
Control Happiness
Control Surprise
Control Fear
Control Sadness
Control Disgust
Control Anger
Patient Happiness
Patient Surprise
Patient Fear
Patient Sadness
Patient Disgust
Patient Anger

Happy:Surprise Surprise:Fear Fear:Sadness Sadness:Disgust Disgust:Anger Anger:Happy